

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2001, 18:42:19 ; Search time 2080.33 Seconds  
(without alignments)  
9549.392 Million cell updates/sec

Title: US-09-494-297-1  
Perfect score: 2274  
Sequence: 1 atgaaaaaaaaggttcc.....ggataagaacaatgactag 2274

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthun10:\*  
45: em\_esthun11:\*  
46: em\_esthun12:\*  
47: em\_esthun13:\*  
48: em\_esthun14:\*  
49: em\_esthun15:\*  
50: em\_esthun16:\*  
51: em\_esthun17:\*  
52: em\_esthun18:\*  
53: em\_esthun19:\*  
54: em\_esthun20:\*  
55: em\_esthun21:\*  
56: em\_esthun22:\*  
57: em\_esthun23:\*  
58: em\_esthun24:\*  
59: em\_esthun25:\*  
60: em\_esthun26:\*  
61: em\_esthun27:\*  
62: em\_esthun28:\*  
63: em\_esthun29:\*  
64: em\_esthun30:\*  
65: em\_esthun31:\*  
66: em\_esthun32:\*  
67: em\_esthun33:\*  
68: em\_esthun34:\*  
69: em\_esthun35:\*  
70: em\_esthun36:\*  
71: em\_esthun37:\*  
72: em\_esthun38:\*  
73: em\_esthun39:\*  
74: em\_esthun40:\*  
75: em\_esthun41:\*  
76: em\_esthun42:\*  
77: em\_esthun43:\*  
78: em\_esthun44:\*  
79: em\_esthun45:\*  
80: em\_esthun46:\*  
81: em\_esthun47:\*  
82: em\_esthun48:\*  
83: em\_esthun49:\*  
84: em\_esthun50:\*  
85: em\_esthun51:\*  
86: em\_esthun52:\*  
87: em\_esthun53:\*  
88: em\_esthun54:\*  
89: em\_esthun55:\*  
90: em\_esthun56:\*  
91: em\_esthun57:\*  
92: em\_esthun58:\*  
93: em\_esthun59:\*  
94: em\_esthun60:\*  
95: em\_esthun61:\*  
96: em\_esthun62:\*  
97: em\_esthun63:\*  
98: em\_esthun64:\*  
99: em\_esthun65:\*  
100: em\_esthun66:\*  
101: em\_esthun67:\*  
102: em\_esthun68:\*  
103: em\_esthun69:\*  
104: em\_esthun70:\*  
105: em\_esthun71:\*  
106: em\_esthun72:\*  
107: em\_esthun73:\*  
108: em\_esthun74:\*  
109: em\_esthun75:\*  
110: em\_esthun76:\*  
111: em\_esthun77:\*  
112: em\_esthun78:\*  
113: em\_esthun79:\*  
114: em\_esthun80:\*  
115: em\_esthun81:\*  
116: em\_esthun82:\*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: em\_gss\_pln1:\*  
 191: em\_gss\_pln2:\*  
 192: em\_gss\_pro:\*  
 193: em\_gss\_rod1:\*  
 194: em\_gss\_rod2:\*  
 195: em\_gss\_rod3:\*  
 196: em\_gss\_rod4:\*  
 197: em\_gss\_rod5:\*  
 198: em\_gss\_vrt1:\*  
 199: em\_gss\_vrt2:\*  
 200: em\_gss\_vrt3:\*  
 201: gb\_gss1:\*  
 202: gb\_gss2:\*  
 203: gb\_gss3:\*  
 204: gb\_gss4:\*  
 205: gb\_gss5:\*  
 206: gb\_gss6:\*  
 207: gb\_gss7:\*  
 208: gb\_gss8:\*  
 209: gb\_gss9:\*  
 210: gb\_gss10:\*  
 211: gb\_gss11:\*  
 212: gb\_gss12:\*  
 213: gb\_gss13:\*  
 214: gb\_gss14:\*  
 215: gb\_gss15:\*  
 216: gb\_gss16:\*  
 217: gb\_gss17:\*  
 218: gb\_gss18:\*  
 219: gb\_gss19:\*  
 220: gb\_gss20:\*  
 221: gb\_gss21:\*  
 222: gb\_gss22:\*  
 223: gb\_gss23:\*  
 224: gb\_gss24:\*  
 225: gb\_gss25:\*  
 226: gb\_gss26:\*  
 227: gb\_gss27:\*  
 228: gb\_gss28:\*  
 229: gb\_gss29:\*  
 230: gb\_gss30:\*  
 231: gb\_gss31:\*  
 232: gb\_gss32:\*  
 233: gb\_gss33:\*  
 234: gb\_gss34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	70.2	3.1	1101	229	AL063921 Drosophila
2	61.4	2.7	1101	229	AL069706 Drosophila
3	61	2.7	1201	229	AL063396 Drosophila
4	60.2	2.6	823	225	AZ687380 EMTL056TF
5	60.2	2.6	1092	230	AZ687380 EMTL056TF
6	60.2	2.6	1101	229	AL175696 Tetradodon
7	60	2.6	1101	229	AL101595 Drosophila
8	59.2	2.6	1032	230	AL06493 Drosophila
9	58.6	2.6	1200	229	AL206746 Tetradodon
10	58	2.6	668	105	AL106578 Drosophila
11	57.8	2.5	891	105	AL514901 AL514901
12	57.4	2.5	652	231	AZ683582 EMTL477R
13	57.4	2.5	1101	229	AL244551 Tetradodon
14	56.8	2.5	1101	229	AL063632 Drosophila
15	56.4	2.5	928	146	AL077628 Drosophila
16	56.2	2.5	881	223	BF274633 GA_EB002
17	56	2.5	1101	229	AZ547573 EMTL056TF
18	55.8	2.5	611	201	AL069440 Drosophila
19					AA549883 0956m3 gm

19	55.8	2.5	877	105	AL514453	AL514453
20	55.8	2.5	884	225	AZ673406	AZ673406
21	55.8	2.5	1101	229	CNS0039L	AL063921
22	55.6	2.4	900	229	CNS01534	AL105514
23	55.4	2.4	636	202	AO159098	AL105909
24	55.2	2.4	828	229	CNS0117X	AL100719
25	55.2	2.4	1025	229	CNS014J2	AL104216
26	54.8	2.4	908	223	AZ548467	AZ548467
27	54.8	2.4	1001	229	CNS0155H	AL105023
28	54.4	2.4	750	229	CNS0110P	AL100303
29	54.2	2.4	734	229	CNS010MP	AL099163
30	54	2.4	767	229	CNS00A0X	AL055924
31	54	2.4	928	229	CNS00DKY	AL071865
32	54	2.4	982	205	AO325799	AO325799
33	53.8	2.4	843	225	AZ351618	AZ351618
34	53.6	2.4	640	18	AL258374	AL258374
35	53.6	2.4	877	223	AZ4545974	AZ4545974
36	53.4	2.3	699	230	CNS03C21	AL237267
37	53.4	2.3	757	105	AL514421	AL514421
38	53.4	2.3	928	229	CNS00DKY	AL071865
39	53.4	2.3	1101	229	CNS00ETW	AL069847
40	53.2	2.3	994	231	CNS04NOJ	AL298972
41	53	2.3	834	227	B12387	B12387
42	53	2.3	865	225	AZ674559	AZ674559
43	52.8	2.3	950	225	AZ672116	AZ672116
44	52.8	2.3	761	230	CNS02MEX	AL204019
45	52.8	2.3	905	223	AZ550256	AZ550256

ALIGNMENTS

RESULT 1  
LOCUS CNS0039G 1101 bp DNA GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778  
VERSION AL063921.1  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr)  
JOURNAL - Web : www.genoscope.cns.fr

COMMENT  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone.lib="RPCI-98"  
/clone="BACR08K10"

BASE COUNT 201 a 64 c 131 g 202 t 503 others  
ORIGIN

Query Match 3.1%; Score 70.2; DB 229; Length 1101;  
Best local similarity 17.5%; Pred. No. 1.8e-05;  
Matches 121; Conservative 286; Mismatches 283; Indels 1; Gaps 1;

Qy	107	tgatcttcgcttgtaagtaacatggtgctgtaagacgtgttggcttagtaaat	166
Db	1091	WBRDTRKDDMDWTWKMTWKBRADBRMAGDADBRMWDGAGTWTATWMMWMTATWD	1032
Qy	167	cctcgagcccaaacgaataatccagatccagatccggaatacagatgtagatag	226
Db	1031	TWMDKMMWATKATDTATWMTARADBRMAGDRGARRDRRAADADGARRDR	972
Qy	227	aatcttaagtaagagccatcatatataaacaagtttagtagtaacagatgaag	286
Db	971	KDKRDKDDDDKGGKKKAKAKAKATKMMDDMDKMDKDKAKDKADDGAGDK	912
Qy	287	ttaacttaagagtaagtaagttatcaagttatgcttaattaaagaacatlc	346
Db	911	DDGKAKDADDPTDGTKDDDDKMDKMDKAGTGDATWMAATDMMWMAADMMWT	852
Qy	347	ctctcgatccagatagtagtcttaaaagtgatataaaacaatgtagatctaca	406
Db	851	DAAADDMWADDRWDAMWMDAMWAGTARDDMDGARRGARRRRDRADDKRD	792
Qy	407	aattgaagattatgtagagccctgaattacaggagatgagctaaatcagaattac	466
Db	791	AADDRDAAATWTTTTRTTRDTRDMDKMTDTWTRMAADTWDDDDDDKRRGTR	732
Qy	467	gagcttgatataatgagacatccacaatgccaatgtaattatgtaagagcttg	526
Db	731	RMKRRMKRRDRRMDADADADARDDRRRGDAGAGGKTKGRRRRRRDATTMDRTA	672
Qy	527	ccttgatgctatcagagttacacaagagcggtatgtagtattcctgaataatgc	586
Db	671	WMAADAMWTTTDTDDMDKRRRRRGARRRRRT-TRAAWMTWTKAMDKMWT	613
Qy	587	tttcttaatccagatgaagtttaaaaggagtcgaagtaactggttagtacttc	646
Db	612	DKRDRMAADTWTDARADDMKARARARARARARARARARARARARARARAR	553
Qy	647	aattactcttgatgcgcgaacttgtaagcaactgtagtccgaattggaactaa	706
Db	552	AAMWMAWMAATTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT	493
Qy	707	tgccaacaagaatccgagatattcaagtaagatttttagtctgaggaagaagg	766
Db	492	AAAAAAMWATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT	433
Qy	767	ataaatataataagatatacaaatcttt	797
Db	432	WMAATTAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT	402

RESULT 2  
LOCUS CNS00EVL 1101 bp DNA GSS 04-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL069706  
VERSION AL069706.1 GI:4949849  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)



D  
918 AAAAAAAAAAATTTTTTTTAAAAAAAATAAAATTAATAAAATTAAATAAATAAANA 859

OY 1736 gaactcagtcgacccagaagtatttagtcatatctcgtaagataaaaaagaag 1795

Db 858 AAAAAAAAAAAAAAAAAAAAAAAWTTTWATTTAAWTAAATTTTAAAAAAAATTTAAAT 799

OY 1796 ttataaccgtaaccaabatttaacatctgagaaaaggctgcgtttgaactgtgcga 1855

Db 798 TTTTNTNNAAAAATNAANAAAWMAAMMAAAWMTAAATTAATAAATTCATNCAATAA 739

OY 1856 gaactaaagattccatttgaaattgatatgataaaataaatcaagaagattgcttc 1915

Db 738 ACAAAAAAATAATTTWAATTAATAAAAAAAAAAAAAAAAAAAAAAANNITTTTTTAACAM 679

OY 1916 aaactgttaaacsagataaaacaacctcgaaattaaagatgtaagcaaccaattaat 1975

Db 678 MAMMAMMMNMNMCCAAAAAAMMAAVMAAACMAMMTTAAAGCNTTAAAAAACCAAMAAAAA 619

OY 1976 taaacatgggggaagttaacactcaaggcttacccaagaagttatcttaacctgcga 2035

Db 618 MAMMAATTAAMAAAAATCTAKRNDKKKMAADKDNAKASMBKRMAATAATWACTGCGC 559

OY 2036 aagaa 2041

Db 558 CABAMA 553

RESULT 4

AZ687380 823 bp DNA GSS 14-DEC-2000

LOCUS ENTLOS6TF Entamoeba histolytica Sheared DNA Entamoeba histolytica

DEFINITION genomic, DNA sequence.

ACCSSION AZ687380

VERSION A2687380.1 GI:11824526

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica.

Eukaryota; Eutrombicidae; Entamoeba.

1 (bases 1 to 823)

Lofthus,B., Van Aken,S. and Fraser,C.

Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library

Unpublished (2000)

Contact: Brendan J Lofthus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel.: 301 838 0200

Fax: 301 838 0208

Email: b.j.lofthus@igrr.org

Claones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Sed primer: ML3-Forward

Class: Shotgun

High quality sequence start: 22

High quality sequence stop: 820.

Location/Qualifiers

1..823

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_id="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOSt1; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (IGRR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, V.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Query Match	2.6%	Score 60.2	DB 225	Length 823
Best Local Similarity	45.9%	Pred. No. 0.003		
Matches 286	Conservative 0	Mismatches 328	Indels 9	Gaps 2
1583	aaccaaagactatcagttgtttgagacatbaatgatagtccttaagcattgctaataa	1642		
67	AAACACATTTGAAACCTTGCGTGAACGAAAGAAAGCTTCTCTCTCAATTCCTCTACTA	126		
1643	tcctttagaatacgcctcaagatagtaatccctccacagctaacctgaccttattcttta	1702		
127	AGAAAGATCTGATGAAAGAGATCTTAAAGTCACAAATCTTGCAANTGAGAAATTAAC	186		
1703	ttcgcgaatacaataataatcaatcctctattggaacctcagtgagcatccagaagattag	1762		
187	CTAAGAAATCTTACTAAAAACAAGCGAATTTGGAAAGAGATTAAGTGAATTAACAAAG	246		
1763	ttgatattatctgtagaagaataaaaagaagtataccttlaactatataattacaat	1822		
247	ATTATGAGATTTTGAAGATGATTAADA-----TAAAGTTGAAGCTATTTGAGCAATG	300		
1823	tgaagaaacggtgacgtggttagctgtgtgacagaactaaagattccatttgaatg	1882		
301	CTCAAGAAAGATTAAGAAAGCTTGATGATGAATTAATCTTAAGACAGATGTTTCACAT	360		
1883	aattaanaataataagcaagaattgcttcccaacatgcttaaacagataaacaacc	1942		
361	ATCTTCAAAAAACAAAAAGAGATATGAATCAACAAATGCTTAAATGCAAGAGAAAAAG	420		
1943	tcgaattt---aaagatggtgaagaacacatcaatttaaacatgvggaaagttaaac	1999		
421	AAGCAATTTGGAATGATGTTTAAAAATTAAGAAAAAGCAATTTAAACAAAGAAATTTAG	480		
2000	ttcaaggtttaccagaaggttattcttaccctgttcaagaacagatctcgaagctata	2059		
481	TTCAAAAGCTTACAAAGAAACCTTGATGAAGAACGAAGTTGAAAMAAAGATGCACAGAA	540		
2060	aggttaagttatagcccaagaagtgcgaatgcacagttccaacacaggaataca	2119		
541	AGAAAGAAAGAAATTTGAAAGGAATATAAGCACTTCACAAAGAAAAAATAATGCGAA	600		
2120	gtgtagtgacactgtcctttgaaataataaagagcgtgttccctcagagttgac	2179		
601	GTAATGAAGATTTCTACGAAAGATTAAGACAAACCTTGAAATTAATTTGAAAGATTA	660		
2180	aaaagatcaatggctactagct	2202		
661	AAAAGAAACCTTGATGATATGACT	683		

BASE COUNT 400 a 81 c 163 g 179 t

Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).



RESULT	7
CNS00EPO/c	CNS00EPO 1101 bp DNA GSS 04-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION	BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL069493
VERSION	AL069493.1 GI:4949636
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 1101)
TITLE	Genoscope.
JOURNAL	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .
FEATURES	Location/Qualifiers
source	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_1id="RPCI-98"
	/clone="BACR29P01"
	/note="end : 77"
BASE COUNT	289 a 155 c 118 g 307 t 232 others
ORIGIN	
Query Match	2.6%; Score 60; DB 229; Length 1101;
Best Local Similarity	33.0%; Pred. No. 0.0034;
Matches 190; Conservative	94; Mismatches 291; Indels 1; Gaps 1;
Db	1030 DMRGTGTATRRRTGCTTAKTKWATBARMMWTADGATGCTGTGTTTTRTTRTKDKAMA 971
Y	1561 taactaagaactactacgtgtcttcttggaagaatgatagatcttagcagtgctaa 1640
D	1030 DMRGTGTATRRRTGCTTAKTKWATBARMMWTADGATGCTGTGTTTTRTTRTKDKAMA 971
Y	1641 aactcttgtagaataacgtcacaagaatgaatccctccacaagcctaactgacctgattctt 1700
D	970 WMTTATTTTWTMTWMTGWTWTTAKDTRTTDWTWTTADTKTAATTGAGTGAAMAATTWTMATT 911
Y	1701 taticgataacataataataatcaatctctatctatggaactcagtgcatccagaagatt 1760
D	910 WATAKATTTTTRAGARTKTKTGTRRTTRDTTDDKAKACTTTTGGATRGAGAKAGATTMTA 851
Y	1761 agtgcatacttctgcgtatgagaataaaaaagaagatctatcctcgttaactcataattaac 1820
D	850 TGTAATATATDTAGWMTATAAATTTAAAAARAAMWTTTWTATATAMAAATTTRTTAANAA 791
Y	1821 attgagaaaaacggtgcgtcttgctgcgcggtgacgaactaaagttccattccattgaat 1880
D	790 ARMDTATTAATAAATAATATATATTTTAAATAAAMWATTTTATATATATTTTAAATTTTMTT 731

Oy	1881	tgaattaaataatataagaagaaatctgttcctcaactgtaataacagttaaaca	1940
Db	730	TTTTCGATTAAMWMAAAAAAACA-M-TAATTARATATAATTAAWMAWTAATATATMTATTTTAA	672
Oy	1941	cctcgaattaaagatgytaagaaccaacttaattaaacaitygggaanglltaacct	2000
Db	671	TTTTTATWMTAADWMTTATTAATAATATMTATTAATTAATAATAATMAWMAAAAATTT	612
Oy	2001	tcaaagtttccagaaggttatcttacccttgccaagaacaagtlctgaagctataa	2060
Db	611	TTTAADTTTAAATTAAWMAWTTAAATAAAMAATTAATGCAGAAATTTTWTATATTATMTTW	552
Oy	2061	ggttaaatataatagccaagaagtagcaaatgctaacagttlcaaaaacaaggatatcaag	2120
Db	551	TTKTAATTAATAAATPTTMAWTTAATTAATWTAATATTTTWTTTTTTMMTTTTATTPAAATN	492
Oy	2121	tgatgagacactgcgttttgaataatataagaagcc	2156
Db	491	AACCAANNAANNNGCTCCGCCCNFTTTTAAACC	456

RESULT	8
LOCUS	CNS0201P
DEFINITION	CNS0201P 1032 bp DNA GSS 14-MAY-2000 Tetradon nigroviridis genome survey sequence PUC-Orig end of clone 15N05 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION	AL206746.1 GI:7865565
VERSION	AL206746
KEYWORDS	GSS; genome survey sequence. Tetradon nigroviridis.
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetradon.
REFERENCE	1 (bases 1 to 1032) Roest-Crollius,H., Jalllon,O., Dasilva,C., Fitzames,C., Fisher,C., Bonneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
TITLE	Unpublished
JOURNAL	2 (bases 1 to 1032)
REFERENCE	RoeSt-CrOllIus,H., Jalllon,O., DasIlva,C., Bouneau,L., Fisher,C., BernOt,A., FItzameS,C., WInckeR,P., BrOtteR,P., QueLler,F., SaUrIn,W. and WeIssenbaCh,J. Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
AUTHORS	Unpublished
TITLE	3 (bases 1 to 1032)
JOURNAL	Genoscope.
REFERENCE	Direct Submission
AUTHORS	Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetradon">http://www.genoscope.cns.fr/tetradon</a> .
FEATURES	Location/Qualifiers
SOURCE	1..1032 /organism="Tetradon nigroviridis" /db_xref="taxon:99883" /clone="15N05" /clone_1bp="G" /note="Genoscope sequence ID : CXKG153CG03SP1-end : PUC-Orig"
BASE COUNT	489 a 103 c 100 g 284 t 56 others
ORIGIN	
Query Match	2.6% Score 59.2; DB 230; Length 1032;
Best Local Similarity	41.5%; Pred. No. 0.0051;





Best Local Similarity 36.4%; Pred. No. 0.0089;  
Matches 20; Conservative 67; Mismatches 287; Indels 4; Gaps 2;

```

QY 1569 attagaataagataaactaaagatacctggttttgagacatga-ataatagtaact 1627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 ATGAATAAAGAAAAAATGACCTGAATTTGAACTGGAAGCAGCAATCMTTAGGTT 557
QY 1628 tagcagtgctaaatccctgtagaatacgtcaagataagtaatccctccacagtaactg 1687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 TATATAAGGGGCTGCTGTGCTGAAAGTAAAMAAATGTTTATTAATAAARATCTGTGG 497
QY 1688 acctgtgcttcttccatccgaataataatcaatccctctatcggaactcagtcg 1747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 RTTWTGTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 437
QY 1748 atccagaagattagtgatatactgctagtgagataaaagaagtataactgtaa 1807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 NNNAATTTNNNNNNNNNAAGCTTWARRRMAAATAAAMAAARRAAMAAATAAAMAA 377
QY 1808 ctcaatacttaacatcgagaaaacggtgactggttagctgtgacagaactaagatt 1867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 WAAAAAATAAGRCWAAATAATARRMMMTAAGGACCTTTAAMAAAAAATAAAMAA 317
QY 1868 tccatttgaattgaattaaataataagaagaattgcttcccaactgttaaa 1927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 AATCWRRTATTAAMMTTWTMAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 257
QY 1928 cagataaacaacactcgaatttaagaatgttaagaacacatatttaaacatg 1987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 197
QY 1988 aaagtttaacacatcgatggttaccagaaggttaacttaccgttcaagaagaact 2047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 WAAAAATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 137
QY 2048 ctgagaagcttaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaag 2107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 TTCCAAATTTTCA---WAAATTTATRAAAAAAATAAATAAATAAATAAATAAATAA 80
QY 2108 caggaataacaagtcagtcagaca 2130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 AAAAAATWATACMTWTWAMCAAMWA 57

RESULT 11
AZ683582 891 bp DNA GSS 14-DEC-2000
LOCUS ENTXK477R Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION AF683582
VERSION AF683582.1 GI:11820728
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 891)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9112 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 694.

```

## FEATURES

source location/Qualifiers

1..891 /organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1, Site-1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome sequencing: A Practical Approach, eds. M. Vaubin and B. Barel, Oxford University Press, 1999)."

BASE COUNT 400 a 55 c 291 g 145 t

ORIGIN

Query Match 2.5%; Score 57.8; DB 225; Length 891;  
Best Local Similarity 43.1%; Pred. No. 0.01;  
Matches 278; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

```

QY 1547 attattcactgtagtgcgaattagaataagaataaactaaagactatcgtttg 1606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39 ATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 98
QY 1607 gagaatgaaatgatacttagcttagcagtgctaaatccctgtgaaatagcctcaagata 1666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 ATGATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 158
QY 1667 gtaatccccaagctaacctgacttattcttattccgaataaataataatcaat 1726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 ATGATGATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 218
QY 1727 ctctattggaactcagtggtcaccagaagaattagtgatatattcgtatgagaata 1786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 278
QY 1787 aaaaagaagttatacctgttaactataaatttaacattgagaagaacgggtgactgttag 1846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 AAGAAAGCAGATGAAGAAGACGATGMAAGACGATGAAGAAGACGATGATG 338
QY 1847 ctggtgacagaactaaagattccatttgaattgaattgaataaataataagaagaat 1906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 AAGAAAGCAGATGATGAAGAAGACGATGATGATGATGATGATGATGATGATGATG 398
QY 1907 tgccttcacaactgtttaaacaagataaacaaccccgaaatttaagaatgtaagcaaa 1966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGA 458
QY 1967 ccattaatttaaacaatggtggaagtttaaacactcaaggtttaaccagaaggttaacct 2026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGA 518
QY 2027 acctgtcagaagaacagattcgaagctataaggttaagtttaagccaagaagtag 2086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGA 578
QY 2087 caaatgttaagttcaaaaacaggaataaacaagtgtgagacactgtctttgaaata 2146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGA 638
QY 2147 ataagaagcctgtgtctccacagaggttgatcaaaagatcaatg 2191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 AAGACGATGAAGAAGATGATGAAGAAGACGATGAAGAAGACGATGAAGAAGATG 683

RESULT 12

```

```

CNS03HOU      652 bp      DNA      GSS      17-MAY-2000
LOCUS          Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION    027K10 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION     AL244551.1 GI:7965563
VERSION       GSS; genome survey sequence.
KEYWORDS      Tetraodon nigroviridis.
SOURCE        Tetraodon nigroviridis.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
              Holacanthopterygii; Acanthopterygii; Percomorpha;
              Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE     Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
AUTHORS       Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE         Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
              Unpublished
              2 (bases 1 to 652)
              Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.
              Human gene number estimate provided by genome wide analysis using
              Tetraodon nigroviridis DNA sequence
              Unpublished
              3 (bases 1 to 652)
              Genoscope.
              Direct Submission
              Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
              This sequence is a single read and was generated as part of a large
              scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/Tetraodon.
              Location/Qualifiers
                1. 652
                  /organism="Tetraodon nigroviridis"
                  /db_xref="taxon:99883"
                  /clone_lib="G"
                  /note="Genoscope sequence ID : COBG027BF05UP1-end : T7"
BASE COUNT    385 a      51 c      41 g      73 t      102 others
ORIGIN
Query Match      2.5%; Score 57.4; DB 231; Length 652;
Best Local Similarity 38.4%; Pred. No. 0.012;
Matches 178; Conservative 4; Mismatches 282; Indels 0; Gaps 0;

QY 1535 agtagcaatataatttcactcgatgtagtgcgaatagataagataaact 1594
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 173 AANTNNANNGTAAATATTCATANANGAAMNANAAAAAANAAAAANANA 232
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1595 atcatggttcggagatgagatgacttagcagtcgtaacatcctctgtaga 1654
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 AANNAAAAAANGAANAANANANANANANANANANANANANANAGTTAAAG 292
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1655 acgctcaagatagatcccccacagctgaactgattctctatccgaataca 1714
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 293 NNCANAAAAAANANANANANANANANANANANANANANANANANANANAA 352
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1715 ataatatcaatcctctatggaactcagtcgacccagaagattgtagataatc 1774
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 AAAAAAAAMWTATTTTAAANANANANANANANANANANANANANANANAN 412
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1775 gtatggaagataaagaaggtatatacctgtaactcataattaacttgagaaac 1834
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 413 ANAAAAAAGAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 472
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1835 tgactgttagctgtgacagaaattccatttgaaatgaaataaataaata 1894

```

```

DB 473 AAAAAAANANANANANANANANANANANANANANANANANANANANANAN 532
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1895 ataagcaagatgcttctccaacgcttaaacagataaacaacccgaaattaa 1954
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 533 ANAAAAAANANANANANANANANANANANANANANANANANANANANANAN 592
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1955 atgtaagcaaccatttaataacatgagggaagtttaaca 1998
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 593 AAAAAAANANANANANANANANANANANANANANANANANANANANANAN 636
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
CNS000B8/c     1101 bp      DNA      GSS      03-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION    BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION     AL063632.1 GI:4938680
VERSION       GSS.
KEYWORDS      fruit fly.
SOURCE        Drosophila melanogaster
              Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 1101)
              Genoscope.
              Direct Submission
              Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mamoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              p1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
              Location/Qualifiers
                1. 1101
                  /organism="Drosophila melanogaster"
                  /db_xref="taxon:7227"
                  /clone_lib="RPCI-98"
                  /clone="BACR01A24"
                  /note="end : TET3"
BASE COUNT    228 a      114 c      110 g      512 t      137 others
ORIGIN
Query Match      2.5%; Score 57.4; DB 229; Length 1101;
Best Local Similarity 35.2%; Pred. No. 0.013;
Matches 175; Conservative 65; Mismatches 257; Indels 0; Gaps 0;

QY 1694 attcttctcgaataaataataatcaatcctctatggaactcagtcgacccag 1753
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1079 ATTAAATATTWWAAAAAANAAAAAATTAATAAAATWATTTWWAAATTAATA 1020
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1754 aagattagtgatattatcgatggaagataaagaagttataactgtaacctata 1813
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1019 AATTWTAATTAATAAAATWTAATAATATATTAAATTAATAAAATWTAATA 960
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1814 atttaacattgagaaacgggtgactggttagctggtgacgaactaagaattccatt 1873
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 959 AAAAAAANANANANANANANANANANANANANANANANANANANANANAN 900

```



## FEATURES

Location/Qualifiers  
1..928

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db\_xref="taxon:29729"

/clone="GA\_Eb0021j19f"

/clone.lib="Gossypium arboreum 7-10 dpa fiber library"

/issue\_type="Fibers Isolated from bolls harvested 7-10 dpa"

/lab\_host="E. coli"

/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 500 a 85 c 80 g 248 t 15 others

ORIGIN

Query Match 2.5%; Score 56.4; DB 146; Length 928;

Best Local Similarity 47.1%; Pred. No. 0.021;

Matches 171; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
QY 1770 tattcgtaiggaagataaaaaagaagttacctgtaactcataatttaacattgagaa 1829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 TATTAAGATTCAAAATACAAAATAATTAGCCGAAAAATTAAGATAAAATGCAAA 465

QY 1830 aacggtgactggttagctggtgacagaactaaagattccatttgaattgaataa 1889
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 466 ACACGTAGATTACCCCAATTAAATCAAAATATAAAAAATCAGCTATACAAAATAANAAA 525

QY 1890 aaataataagaagaattgcttctcaaaactgttaaaacagataaaacaacctggaatt 1949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 AAAAAAAAAAAAAAAAAAACACAAAATACTAAACAAAAAAATTAATTAACATAAAAAA 585

QY 1950 taaagtggtaaagcaaccttaattaaacatgggaaagttaaccttaaggtt 2009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 TAAAAAGAAATTAATAATTAATAAGCAAAAAAGAAAAATTAATAATTAATTAATTAAT 645

QY 2010 accagaaggttattcttaccctgtcaagaagaacagattctgaagctataaggttaaagt 2069
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 646 TAAAAATATATTAATAAAAAACGAAAAACAAAGAAATTAATAAGACAAAAAATAGTAATA 705

QY 2070 taatagccaagaagtagcaaatgtctacagttcaaaaaacagaataacaagtgtatgagac 2129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 AAATATATAAAATATTAATAATAACAAATAAATAATTAATAAAGAGAAATATATTAAT 765

QY 2130 act 2132
    | |
Db 766 AAT 768
```

Search completed: June 6, 2001, 20:03:01  
Job time: 4842 sec